

ВИКОРИСТАННЯ ОНТОЛОГІЧНОГО ПІДХОДУ ДЛЯ ПОБУДОВИ МОДЕЛЕЙ ПОШУКУ ТА КЛАСИФІКАЦІЇ ПОСЛУГ ТА ДОКУМЕНТІВ

Мороз Б.І., Костенко Д.Є., Костенко В.В., Університет митної справи та фінансів, м. Дніпропетровськ

На сьогоднішній день інформаційний пошук швидко стає основною формою доступу населення до інформації у суспільстві. У різних містах України інформаційні центри господарства надають велику кількість послуг. Користувачу у таких випадках потрібні досконалі засоби пошуку інформації. У даному випадку річ може йти про швидкий доступ до потрібних послуг.

Крім цього, в процесі експлуатації системи може виникнути необхідність зміни переліку оброблюваних послуг. Це пов'язано з тим, що з кожним роком збільшується об'єм доступних користувачу масивів текстової інформації. Це сприяє більшій актуалізації задачі пошуку необхідної користувачу інформації в таких масивах. Для вирішення подібних задач дуже часто використовуються елементарні тематичні класифікатори, рубрикатори. Але не завжди такий результат пошуку може відповідати поставленим завданням.

До інформаційного пошуку можна також віднести деякі завдання, які не підпадають під базове визначення. Коли ми говоримо «неструктуровані дані», ми маємо на увазі дані, які не мають ясної, семантично очевидної й легко реалізованої на комп'ютері структури. Це – протилежність структурованим даним, канонічним прикладом яких є реляційні бази даних (які звичайно використовуються підприємствами для зберігання реєстрів послуг й персональних даних співробітників). У реальності, зовсім «неструктурованих даних» практично не існує. Наприклад, звичайні текстові дані мають сховану структуру, характерну для природних мов.

Аналіз темпів зростання кількості документів, послуг та методів їх обробки, зокрема, у сфері міського та регіонального господарства, наочно показує, що традиційні механізми роботи з інформацією не задовольняють потребам сучасного користувача. Стратегічне значення цього моменту полягає в тому, що послуг(та супровідних до них документів) може бути величезна кількість, яка зростає і зростає.

До інформаційного пошуку відносяться й такі завдання, як навігація користувачів по колекції документів і послуг, їх фільтрація, а також подальша обробка. Якщо є набір послуг або документів, то виникає завдання кластеризації, що полягає у визначенні найкращої сукупності за змістом. Це нагадує розміщення книг на полицях по темах.

Процес пошуку може відбуватися в досить великій базі даних. Буває, що навіть на конкретно сформований запит(саме користувачами), може бути досить неточний результат пошуку.

Завжди існує дилема: чому віддати перевагу – потужним обчислювальним процедурам, що опираються на відносно невеликі словникові системи з багатою граматичною й семантичною інформацією, або потужному декларатив-

ному компоненту при відносно простих процедурних засобах? Питання залишається відкритим.

Навіть у величезних масивах інформації повинні працювати пошукові системи, які забезпечували б користувачу результат – швидкий і точний. Зазвичай використовується ряд аналітичних підходів до подання інформації для забезпечення її наступного пошуку. Деякі базуються на теорії множин, інші – на елементах векторної алгебри. Обидва підходи ефективно реалізуються в умовах практики.

Інформація достовірна, якщо вона показує реальне положення справ. Об'єктивна інформація завжди достовірна, але достовірна інформація може бути як об'єктивною, так і суб'єктивною. Достовірна інформація допомагає прийняти нам правильне рішення, зробити правильний вибір без зайвих затрат часу та коштів.

Моделі пошуку визначають міру відповідності між запитом та результатом. Взагалі, можна вивести загальну ідею: більший відсоток збігу між запитом та результатом вказує, що результат є більш відповідним до запиту користувача. Поняття «збіг» є багатоаспектним: включається термінологічна база, ключові поняття, співпадіння предметних областей.

Одним з можливих варіантів вирішення проблем пошуку інформації можна назвати семантичний пошук. Він надає результат не тільки за заданими словами з запиту, але й за еквівалентними за сенсом.

Для ефективного семантичного пошуку необхідна інформація про предметну область, про властиві їй поняття та відношення між ними, а також про обмеження, які існують між відношеннями. Таку інформацію прийнято називати онтологією. Онтологічна модель може бути використана як для повнотекстового пошуку, так і для окремої класифікації. Для побудови онтології потрібно формальне декларативне подання чітко організованих конструкцій, які містять у собі словник термінів тематичної області, опис визначень цих термінів, існуючі взаємозв'язки між ними, і взагалі – теоретично можливі й неможливі взаємозв'язки.

Важливість підходу, пов'язаного з онтологіями, обумовлена також тим, що знання, яке не описане і не тиражоване, в кінцевому рахунку стає застарілим і непотрібним. Навпаки, знання, яке поширюється, є генератором нових знань. Особливу увагу при роботі з онтологіями треба приділяти засобам пошуку по тексту на основі шаблонів – регулярним виразам. Шаблон описує закономірність, якій повинні підкорятися шукані послідовності символів у тексті.

Проблема полягає в тому, щоб зробити пошук послуг або документів динамічним і зручним для користувача. Для будь-якого типу запиту, що виникає в практичній діяльності, повинні бути знайдені адекватні знання в інформаційному просторі. При цьому мова для формулювання пошукової вимоги не повинна бути занадто складною.